(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
9 August 2001 (09.08.2001)

PCT

(10) International Publication Number
WO 01/57278 A2

(51) International Patent Classification[7]: C12Q 1/68, G06F 19/00, C07K 14/47

(21) International Application Number: PCT/US01/00670

(22) International Filing Date: 30 January 2001 (30.01.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:

60/180,312	4 February 2000 (04.02.2000)	US
60/207,456	26 May 2000 (26.05.2000)	US
09/608,408	30 June 2000 (30.06.2000)	US
09/632,366	3 August 2000 (03.08.2000)	US
60/234,687	21 September 2000 (21.09.2000)	US
60/236,359	27 September 2000 (27.09.2000)	US
0024263.6	4 October 2000 (04.10.2000)	GB

(71) Applicant *(for all designated States except US)*: **MOLECULAR DYNAMICS, INC.** [—/US]; 928 East Arques Avenue, Sunnyvale, CA 94086 (US).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **PENN, Sharron, G.** [GB/US]; 617 South Delaware Street, San Mateo, CA 94402 (US). **HANZEL, David, K.** [US/US]; 988 Loma Verde Avenue, Palo Alto, CA 94303 (US). **CHEN, Wensheng** [CN/US]; 210 Easy Street #25, Mountain View, CA 94043 (US). **RANK, David, R.** [US/US]; 117 El Dorado Commons, Fremont, CA 94539 (US).

(74) Agent: **RONNING, Royal, N., Jr.**; Amersham Pharmacia Biotech, Inc., 800 Centennial Avenue, Piscataway, NJ 08855 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— *without international search report and to be republished upon receipt of that report*

— *entirely in electronic form (except for this front page) and available upon request from the International Bureau*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 01/57278 A2

(54) Title: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR ANALYSIS OF GENE EXPRESSION IN HUMAN HELA CELLS OR OTHER HUMAN CERVICAL EPITHELIAL CELLS

(57) Abstract: A single exon nucleic acid microarray comprising a plurality of single exon nucleic acid probes for measuring gene expression in a sample derived from human HeLa cells is described. Also described are single exon nucleic acid probes expressed in the HeLa cells and their use in methods for detecting gene expression.

HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR ANALYSIS OF GENE EXPRESSION IN HUMAN HELA CELLS OR OTHER HUMAN CERVICAL EPITHELIAL CELLS

CROSS REFERENCE TO RELATED APPLICATIONS

The present application is a continuation-in-part of U.S. patent application serial nos. 09/632,366, filed August 3, 2000 and 09/608,408, filed June 30, 2000; claims the benefit under 35 U.S.C. s 119(e) of U.S.provisional patent application serial nos. 60/236,359, filed September 27, 2000, 60/234,687, filed September 21, 2000, 60/207,456, filed May 26, 2000, and 60/180,312, filed February 4, 2000; and further claims the benefit under 35 U.S.C. s 119(a) of UK patent application no. 0024263.6, filed October 4, 2000, the disclosures of which are incorporated herein by reference in their entireties.

REFERENCE TO SEQUENCE LISTING AND INCORPORATION BY REFERENCE THEREOF

The present application includes a Sequence Listing in electronic format, filed pursuant to PCT Administrative Instructions 801 - 806 on a single CD-R disc, in triplicate, containing a file named pto_HELA.txt, created 24 January 2001, having 18,781,468 bytes. The Sequence Listing contained in said file on said disc is incorporated herein by reference in its entirety.

Field of the Invention

The present invention relates to genome-derived single exon microarrays useful for verifying the expression of regions of genomic DNA predicted to encode protein. In particular, the present invention relates to unique genome-

derived single exon nucleic acid probes expressed in human HeLa cells and single exon nucleic acid microarrays that include such probes.

Background of the Invention

For almost two decades following the invention of general techniques for nucleic acid sequencing, Sanger *et al.*, *Proc. Natl. Acad. Sci. USA* 70(4):1209-13 (1973); Gilbert *et al.*, *Proc. Natl. Acad. Sci. USA* 70(12):3581-4 (1973), these techniques were used principally as tools to further the understanding of proteins – known or suspected – about which a basic foundation of biological knowledge had already been built. In many cases, the cloning effort that preceded sequence identification had been both informed and directed by that antecedent biological understanding.

For example, the cloning of the T cell receptor for antigen was predicated upon its known or suspected cell type-specific expression, by its suspected membrane association, and by the predicted assembly of its gene via T cell-specific somatic recombination. Subsequent sequencing efforts at once confirmed and extended understanding of this family of proteins. Hedrick *et al.*, *Nature* 308(5955):153-8 (1984).

More recently, however, the development of high throughput sequencing methods and devices, in concert with large public and private undertakings to sequence the human and other genomes, has altered this investigational paradigm: today, sequence information often precedes understanding of the basic biology of the encoded protein product.

One of the approaches to large-scale sequencing is predicated upon the proposition that expressed sequences – that is, those accessible through isolation of mRNA – are of greatest initial interest. This "expressed

sequence tag" ("EST") approach has already yielded vast amounts of sequence data (see for example Adams *et al.*, *Science* 252:1651 (1991); Williamson, *Drug Discov. Today* 4:115 (1999)). For nucleic acids sequenced by this approach, often the only biological information that is known *a priori* with any certainty is the likelihood of biologic expression itself. By virtue of the species and tissue from which the mRNA had originally been obtained, most such sequences are also annotated with the identity of the species and at least one tissue in which expression appears likely.

More recently, the pace of genomic sequencing has accelerated dramatically. When genomic DNA serves as the initial substrate for sequencing efforts, expression cannot be presumed; often the only *a priori* biological information about the sequence includes the species and chromosome (and perhaps chromosomal map location) of origin.

With the ever-accelerating pace of sequence accumulation by directed, EST, and genomic sequencing approaches — and in particular, with the accumulation of sequence information from multiple genera, from multiple species within genera, and from multiple individuals within a species — there is an increasing need for methods that rapidly and effectively permit the functions of nucleic sequences to be elucidated. And as such functional information accumulates, there is a further need for methods of storing such functional information in meaningful and useful relationship to the sequence itself; that is, there is an increasing need for means and apparatus for annotating raw sequence data with known or predicted functional information.

Although the increase in the pace of genomic sequencing is due in large part to technological changes in sequencing strategies and instrumentation, Service, *Science* 280:995 (1998); Pennisi, *Science* 283: 1822-1823 (1999),

there is an important functional motivation as well.

While it was understood that the EST approach would rarely be able to yield sequence information about the noncoding portions of the genome, it now also appears the EST approach is capable of capturing only a fraction of a genome's actual expression complexity.

For example, when the *C. elegans* genome was fully sequenced, gene prediction algorithms identified over 19,000 potential genes, of which only 7,000 had been found by EST sequencing. *C. elegans* Sequencing Consortium, *Science* 282:2012 (1998). Analogously, the recently completed sequence of chromosome 2 of *Arabidopsis* predicts over 4000 genes, Lin *et al.*, *Nature*, 402:761 (1999), of which only about 6% had previously been identified via EST sequencing efforts. Although the human genome has the greatest depth of EST coverage, it is still woefully short of surrendering all of its genes. One recent estimate suggests that the human genome contains more than 146,000 genes, which would at this point leave greater than half of the genes undiscovered. It is now predicted that many genes, perhaps 20 to 50%, will only be found by genomic sequencing.

There is, therefore, a need for methods that permit the functional regions of genomic sequence – and most importantly, but not exclusively, regions that function to encode genes – to be identified.

Much of the coding sequence of the human genome is not homologous to known genes, making detection of open reading frames ("ORFs") and predictions of gene function difficult. Computational methods exist for predicting coding regions in eukaryotic genomes. Gene prediction programs such as GRAIL and GRAIL II, Uberbacher *et al.*, *Proc. Natl. Acad. Sci. USA* 88(24):11261-5 (1991); Xu *et al.*, *Genet. Eng.* 16:241-53 (1994); Uberbacher *et al.*, *Methods Enzymol.* 266:259-81 (1996); GENEFINDER, Solovyev *et*

al., *Nucl. Acids. Res.* 22:5156-63 (1994); Solovyev *et al.*, *Ismb* 5:294-302 (1997); and GENESCAN, Burge *et al.*, *J. Mol. Biol.* 268:78-94 (1997), predict many putative genes without known homology or function. Such programs are known, however, to give high false positive rates. Burset *et al.*, *Genomics* 34:353-367 (1996). Using a consensus obtained by a plurality of such programs is known to increase the reliability of calling exons from genomic sequence. Ansari-Lari *et al.*, *Genome Res.* 8(1):29-40 (1998)

Identification of functional genes from genomic data remains, however, an imperfect art. For example, in reporting the full sequence of human chromosome 21, the Chromosome 21 Mapping and Sequencing Consortium reports that prior bioinformatic estimates of human gene number may need to be revised substantially downwards. *Nature* 405:311-199 (2000); Reeves, *Nature* 405:283-284 (2000).

Thus, there is a need for methods and apparatus that permit the functions of the regions identified bioinformatically – and specifically, that permit the expression of regions predicted to encode protein – readily to be confirmed experimentally.

Recently, the development of nucleic acid microarrays has made possible the automated and highly parallel measurement of gene expression. *Reviewed in* Schena (ed.), DNA Microarrays : A Practical Approach (Practical Approach Series), Oxford University Press (1999) (ISBN: 0199637768); *Nature Genet.* 21(1)(suppl):1 - 60 (1999); Schena (ed.), Microarray Biochip: Tools and Technology, Eaton Publishing Company/BioTechniques Books Division (2000) (ISBN: 1881299376).

It is common for microarrays to be derived from cDNA/EST libraries, either from those previously described in the literature, such as those from the I.M.A.G.E. consortium, Lennon *et al.*, *Genomics* 33(1):151-2 (1996), or from the construction of "problem specific" libraries

targeted at a particular biological question, R.S. Thomas *et al.*, *Cancer Res.* (in press). Such microarrays by definition can measure expression only of those genes found in EST libraries, and thus have not been useful as probes for genes discovered solely by genomic sequencing.

The utility of using whole genome nucleic acid microarrays to answer certain biological questions has been demonstrated for the yeast *Saccharomyces cerevisiae*. De Risi *et al.*, *Science* 278:680 (1997). The vast majority of yeast nuclear genes, approximately 95% however, are single exon genes, *i.e.*, lack introns, Lopez *et al.*, *RNA* 5:1135-1137 (1999); Goffeau *et al.*, *Science* 274:563-67 (1996), permitting coding regions more readily to be identified. Whole genome nucleic acid microarrays have not generally been used to probe gene expression from more complex eukaryotic genomes, and in particular from those averaging more than one intron per gene.

Summary of the Invention

The present invention solves these and other problems in the art by providing methods and apparatus for predicting, confirming, and displaying functional information derived from genomic sequence. The present invention also provides apparatus for verifying the expression of putative genes identified within genomic sequence.

In particular, the invention provides novel genome-derived single exon nucleic acid microarrays useful for verifying the expression of putative genes identified within genomic sequence.

The present invention also provides compositions and kits for the ready production of nucleic acids identical in sequence to, or substantially identical in sequence to, probes on the genome-derived single exon

CLAIMS

1. A spatially-addressable set of single exon nucleic acid probes for measuring gene expression in a sample derived from human HeLa cells or other human cervical epithelial cells comprising a plurality single exon nucleic probes, said probes comprising any one of the nucleotide sequences set out in SEQ ID NOs: 1 - 9,290 or a complementary sequence, or a portion of such a sequence.

2. A spatially-addressable set of single exon nucleic acid probes as claimed in claim 1 wherein each of said plurality of probes is separately and addressably amplifiable.

3. A spatially-addressable set of single exon nucleic acid probes as claimed in claim 1 wherein each of said plurality of probes is separately and addressably isolatable from said plurality.

4. A spatially-addressable set of single exon nucleic acid probes as claimed in any of claims 1 to 3 wherein said probes comprise any one of the nucleotide sequences set out in SEQ ID NOS.: 9,291 - 18,392.

5. A spatially-addressable set of single exon nucleic acid probes as claimed in any of claims 1 to 4, wherein each of said plurality of probes is amplifiable using at least one common primer.

6. A spatially-addressable set of single exon nucleic acid probes as claimed in any of claims 1 to 5 wherein the set comprises between 50 - 20,000 single exon nucleic acid probes.

7. A spatially-addressable set of single exon nucleic acid

probes as claimed in any of claims 1 to 6, wherein the average length of the single exon nucleic acid probes is between 200 and 500 bp.

8. A spatially-addressable set of single exon nucleic acid probes as claimed in any of claims 1 to 7, wherein at least 50% of said single exon nucleic acid probes lack prokaryotic and bacteriophage vector sequence.

9. A spatially-addressable set of single exon nucleic acid probes as claimed in any of claims 1 to 8, wherein at least 50% of said single exon nucleic acid probes lack homopolymeric stretches of A or T.

10. A spatially-addressable set of single exon nucleic acid probes as claimed in any of claims 1 - 9 characterised in that said set of probes is addressably disposed upon a substrate.

11. A spatially-addressable set of single exon nucleic acid probes as claimed in claim 10 wherein said substrate is selected from glass, amorphous silicon, crystalline silicon and plastic.

12. A microarray comprising a spatially addressable set of single exon nucleic acid probes as claimed in any of claims 1 - 11.

13. A single exon nucleic acid probe for measuring human gene expression in a sample derived from human HeLa cells or other human cervical epithelial cells comprising a nucleotide sequence as set out in any of SEQ ID NOs.: 1 - 9,290 or a complementary sequence or a fragment thereof wherein said probe hybridizes at high stringency to a nucleic acid molecule expressed in the human HeLa cells or

other human cervical epithelial cells.

14. A single exon nucleic acid probe as claimed in claim 13 comprising a nucleotide sequence as set out in any of SEQ ID NOs.: 9,291 - 18,392 or a complementary sequence or a fragment thereof.

15. A single exon nucleic acid probe for measuring human gene expression in a sample derived from human HeLa cells or other human cervical epithelial cells which is a nucleic acid molecule having a sequence encoding a peptide comprising a peptide sequence as set out in any of SEQ ID NOs.: 18,393 - 26,941, or a complementary sequence or a fragment thereof wherein said probe hybridizes at high stringency to a nucleic acid expressed in the human HeLa cells or other human cervical epithelial cells.

16. A single exon nucleic acid probe as claimed in any one of claims 13 to 15 wherein said single exon nucleic acid probe comprises between 15 and 25 contiguous nucleotides of said SEQ ID NO.

17. A single exon nucleic acid probe as claimed in any one of claims 13 to 15, wherein said probe is between 3 - 25 kb in length.

18. A single exon nucleic acid probe as claimed in any one of claims 13 - 17, wherein said probe is DNA, RNA or PNA.

19. A single exon nucleic acid probe as claimed in any one of claims 13 - 18, wherein said probe is detectably labeled.

20. A single exon nucleic acid probe as claimed in any one of claims 13 - 19, wherein said probe lacks prokaryotic and

bacteriophage vector sequence.

21. A single exon nucleic acid probe as claimed in any one of claims 13 - 20, wherein said probe lacks homopolymeric stretches of A or T.

22. A method of measuring gene expression in a sample derived from human HeLa cells or other human cervical epithelial cells, comprising:

contacting the microarray of claim 12, with a first collection of detectably labeled nucleic acids, said first collection of nucleic acids derived from mRNA of human HeLa cells or other human cervical epithelial cells; and then

measuring the label detectably bound to each probe of said microarray.

23. A method of identifying exons in a eukaryotic genome, comprising:

algorithmically predicting at least one exon from genomic sequence of said eukaryote; and then

detecting specific hybridization of detectably labeled nucleic acids to a single exon probe,

wherein said detectably labeled nucleic acids are derived from mRNA from the HeLa cells or other human cervical epithelial cells of said eukaryote, said probe is a single exon probe having a fragment identical in sequence to, or complementary in sequence to, said predicted exon, said probe is included within a microarray according to claim 12, and said fragment is selectively hybridizable at high stringency.

24. A method of assigning exons to a single gene, comprising:

identifying a plurality of exons from genomic

sequence according to the method of claim 23; and then

measuring the expression of each of said exons in a plurality of tissues and/or cell types using hybridization to single exon microarrays having a probe with said exon,

wherein a common pattern of expression of said exons in said plurality of tissues and/or cell types indicates that the exons should be assigned to a single gene.

25. A nucleic acid sequence as set out in any of SEQ ID NOs: 1 - 18,392 which encodes a peptide.

26. A peptide encoded by a sequence as set out in any of SEQ ID Nos: 1 - 18,392.

27. A peptide comprising a sequence as set out in any of SEQ ID Nos: 18,393 - 26,941.

Table 4

Single Exon Probes Expressed in HELA Cells

Probe SEQ ID NO:	Exon SEQ ID NO:	ORF SEQ ID NO:	Expression Signal	Most Similar (Top) Hit BLAST E Value	Top Hit Acession No.	Top Hit Database Source	Top Hit Descriptor
456	9709	18846	5.87				
899	10134	19296	11.93				
1052	10278		2.08				
1309	10525	19685	19.17				
1593	10806	19982	2.78				
1613	10826	20001	12.21				
1700	10912	20099	1.72				
1721	10933	20116	1.1				
1727	10939	20122	9.25				
1856	11063	20254	1.76				
1947	11151	20353	2.13				
2131	11330	20549	1.94				
2244	11439	20663	2.03				
3149	12384	21517	3.28				
3428	12651	21781	1				
3489	12713	21849	8.63				
3535	12756		0.67				
3637	12858	21977	0.97				
3923	13139		1.02				
4179	13383	22484	1.57				
4248	13451	22542	7.61				
4268	13469	22561	0.64				
4268	13469	22562	0.64				
4330	13531		1.28				
4386	13587	22689	0.79				
4854	14043	23137	1.05				
4899	14087		0.89				
5070	14250	23333	5.18				
5404	14632		5.28				
5488	14714		6.43				
5525	14632		4.42				
5546	14770	24137	3.03				
5669	18062	24282	1.69				
5688	14908	24301	1.91				

```
RESULT 4
CQ073428
LOCUS       CQ073428                 575 bp    DNA     linear   PAT 20-JAN-2004
DEFINITION  Sequence 9228 from Patent WO0157278.
ACCESSION   CQ073428
VERSION     CQ073428.1  GI:41043297
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Euarchontoglires; Primates; Catarrhini;
            Hominidae; Homo.
REFERENCE   1
  AUTHORS   Penn,S.G., Hanzel,D.K., Chen,W. and Rank,D.R.
  TITLE     Human genome-derived single exon nucleic acid probes useful for
            analysis of gene expression in human hela cells or other human
            cervical epithelialcells
  JOURNAL   Patent: WO 0157278-A 9228 09-AUG-2001;
            Aeomica, Inc. (US)
FEATURES             Location/Qualifiers
     source          1. .575
                     /organism="Homo sapiens"
                     /mol_type="unassigned DNA"
                     /db_xref="taxon:9606"
                     /note="MAP TO AP000347.1
                     EXPRESSED IN HELA, SIGNAL = 1.5"
ORIGIN

  Query Match             100.0%;  Score 70;  DB 2;  Length 575;
  Best Local Similarity   100.0%;  Pred. No. 3.6e-12;
  Matches   70;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 AAACAGTCTCAGGGAGGCCCGGCTGCAAGACTGGGTGACACACACAGGGAGTGTGGATCT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        138 AAACAGTCTCAGGGAGGCCCGGCTGCAAGACTGGGTGACACACACAGGGAGTGTGGATCT 197

Qy         61 GGGCCAGTGG 70
              ||||||||||
Db        198 GGGCCAGTGG 207
```

SEQ 6

SEQ 25

```
RESULT 2
AAI19295
ID   AAI19295 standard; DNA; 575 BP.
XX
AC   AAI19295;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #9228 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000670.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human cervical epithelial cells.
XX
PS   Claim 25; SEQ ID NO 9228; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging of
CC   diseases of the cervix, notably cervical cancer. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 575 BP; 126 A; 147 C; 204 G; 98 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 128;  DB 4;  Length 575;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-33;
  Matches  128;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TGAGGGTGCTCGTGCCTGGTTCTTCCTCAGAGGGATGACGGTGAGAACAACGGCAACAGC 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         10 TGAGGGTGCTCGTGCCTGGTTCTTCCTCAGAGGGATGACGGTGAGAACAACGGCAACAGC 69
```

```
Qy          61 TACAGGAAACTGAGCCCTCAGAGGCCCTGTGAGGTAGCTGTGGTTTGCATCACTCTTTAC 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          70 TACAGGAAACTGAGCCCTCAGAGGCCCTGTGAGGTAGCTGTGGTTTGCATCACTCTTTAC 129

Qy         121 AGAAGAGG 128
               ||||||||
Db         130 AGAAGAGG 137
```

```
RESULT 3
ABA64305
ID   ABA64305 standard; DNA; 575 BP.
XX
AC   ABA64305;
XX
DT   01-FEB-2002  (first entry)
XX
DE   Human foetal liver single exon nucleic acid probe #12610.
XX
KW   Human; foetal liver; gene expression; single exon nucleic acid probe; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157277-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000669.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-483447/52.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human fetal liver.
XX
PS   Claim 1; SEQ ID NO 12610; 639pp + Sequence Listing; English.
XX
CC   The invention relates to a single exon nucleic acid probe for measuring
CC   human gene expression in a sample derived from human foetal liver. The
CC   single exon nucleic acid probes may be used for predicting, measuring and
CC   displaying gene expression in samples derived from human fetal liver. The
CC   present sequence is a single exon nucleic acid probe of the invention.
CC   Note: The sequence data for this patent did not form part of the printed
CC   specification, but was obtained in electronic format directly from WIPO
CC   at ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 575 BP; 126 A; 147 C; 204 G; 98 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 128;  DB 4;  Length 575;
  Best Local Similarity   100.0%;  Pred. No. 6.9e-33;
```

SEQ 7

```
RESULT 3
AAI19295
ID   AAI19295 standard; DNA; 575 BP.
XX
AC   AAI19295;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #9228 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000670.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human cervical epithelial cells.
XX
PS   Claim 25; SEQ ID NO 9228; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging of
CC   diseases of the cervix, notably cervical cancer. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 575 BP; 126 A; 147 C; 204 G; 98 T; 0 U; 0 Other;

  Query Match             100.0%;  Score 117;  DB 4;  Length 575;
  Best Local Similarity   100.0%;  Pred. No. 4.2e-23;
  Matches  117;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 TATGAGCACGGTGCCAGGTGGCTCCCGCCACTCCCTGGGGATCCAAGTGCGGGGTGGCTG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        208 TATGAGCACGGTGCCAGGTGGCTCCCGCCACTCCCTGGGGATCCAAGTGCGGGGTGGCTG 267
```

```
Qy          61 GGGTGTAACTGGGGGAGAGGAGGAGAGCCTCACTGTCCCTGTCGCTGACACCTGGCA 117
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         268 GGGTGTAACTGGGGGAGAGGAGGAGAGCCTCACTGTCCCTGTCGCTGACACCTGGCA 324
```

SEQ 8

```
RESULT 3
AAI19295
ID   AAI19295 standard; DNA; 575 BP.
XX
AC   AAI19295;
XX
DT   12-OCT-2001  (first entry)
XX
DE   Probe #9228 for gene expression analysis in human cervical cell sample.
XX
KW   Probe; human; microarray; gene expression; cervical epithelial cell;
KW   cervical cancer; ss.
XX
OS   Homo sapiens.
XX
PN   WO200157278-A2.
XX
PD   09-AUG-2001.
XX
PF   30-JAN-2001; 2001WO-US000670.
XX
PR   04-FEB-2000; 2000US-0180312P.
PR   26-MAY-2000; 2000US-0207456P.
PR   30-JUN-2000; 2000US-00608408.
PR   03-AUG-2000; 2000US-00632366.
PR   21-SEP-2000; 2000US-0234687P.
PR   27-SEP-2000; 2000US-0236359P.
PR   04-OCT-2000; 2000GB-00024263.
XX
PA   (MOLE-) MOLECULAR DYNAMICS INC.
XX
PI   Penn SG,  Hanzel DK,  Chen W,  Rank DR;
XX
DR   WPI; 2001-488901/53.
XX
PT   Human genome-derived single exon nucleic acid probes useful for analyzing
PT   gene expression in human cervical epithelial cells.
XX
PS   Claim 25; SEQ ID NO 9228; 487pp; English.
XX
CC   The present invention relates to human single exon nucleic acid probes
CC   (SENP). The present sequence is one such probe. The SENPs are derived
CC   from human HeLa cells. The SENPs can be used to produce a single exon
CC   microarray, which can be used for measuring human gene expression in a
CC   sample derived from human cervical epithelial cells. By measuring gene
CC   expression, the probes are therefore useful in grading and/or staging of
CC   diseases of the cervix, notably cervical cancer. Note: The sequence data
CC   for this patent did not form part of the printed specification, but was
CC   obtained in electronic format directly from WIPO at
CC   ftp.wipo.int/pub/published_pct_sequences
XX
SQ   Sequence 575 BP; 126 A; 147 C; 204 G; 98 T; 0 U; 0 Other;

Alignment Scores:
Pred. No.:                1.44e-16      Length:          575
Score:                    205.00        Matches:         38
Percent Similarity:       100.0%        Conservative:    0
Best Local Similarity:    100.0%        Mismatches:      0
Query Match:              100.0%        Indels:          0
DB:                       4             Gaps:            0
```

US-10-625-471-8 (1-38) x AAI19295 (1-575)

```
Qy          1 MetSerThrValProGlyGlySerArgHisSerLeuGlyIleGlnValArgGlyGlyTrp 20
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        209 ATGAGCACGGTGCCAGGTGGCTCCCGCCACTCCCTGGGGATCCAAGTGCGGGGTGGCTGG 268

Qy         21 GlyValThrGlyGlyGluGluGluSerLeuThrValProValAlaAspThrTrp 38
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        269 GGTGTAACTGGGGGAGAGGAGGAGAGCCTCACTGTCCCTGTCGCTGACACCTGG 322
```